VALUE IN HEALTH 18 (2015) 841–845

Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/jval

Using the Analytic Hierarchy Process to Derive Health State Utilities from Ordinal Preference Data

Brian P. Reddy, MSc[1,*], Roisin Adams, PhD[1], Cathal Walsh, PhD[1,2], Michael Barry, PhD[1,3], Paul Kind, MPhil[4]

[1]National Centre for Pharmacoeconomics, St. James Hospital, Dublin, Ireland; [2]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland; [3]Health Research Institute (HRI)/Mathematics Applications Consortium for Science and Industry (MACSI), University of Limerick, Limerick, Ireland; [4]Institute of Health Sciences, University of Leeds, Leeds, UK

ABSTRACT

Background: The EuroQol five-dimensional questionnaire is a standardized instrument used in the economic evaluation of health care to measure health state preferences across disease groups. A time trade-off (TTO) approach is commonly used to elicit preferences from the public. However, there are issues regarding how best to measure worse-than-dead states; at present, extreme valuations are rounded up to more acceptable values. TTO elicitation is also cognitively demanding for respondents and is therefore expensive to investigate. **Objectives:** To describe how the analytic hierarchy process approach could be used to generate utilities from the ordinal relationships between the health states instead of the ordinal relationships between health states, allowing potentially useful preference data to be incorporated rather than excluded as they are at present. It was applied to the Measurement and Valuation of Health study data set, measuring health state preferences for the United Kingdom. **Methods:** The analytic hierarchy process approach was explained. Five approaches to structure pairwise comparisons of health state preference were described (two concave, two convex, and one linear). **Results:** All approaches described predicted the rankings of health states well. However, utilities derived followed an unconventional, bunched shape compared with the original Measurement and Valuation of Health TTO study. An approach was identified by optimizing the parameters, minimizing the sum of squared errors between the ordinal “health state ranking” approach and the original TTO-derived utilities. **Conclusions:** This approach outlined offers the potential to convert ordinal preference data into cardinal utilities. It is simpler than TTO studies to carry out and removes the need to directly alter results of the preference ranking exercise.

Keywords: analytic hierarchy process, EQ-5D, health state valuation, utilities.

Background

Over recent decades, economic evaluation of health care issues has become increasingly important in structuring and informing subsequent decisions. Given a limited available budget, the provision of a new drug or other intervention is assumed to carry an opportunity cost, displacing “health” somewhere else in the system. This must be compared against the benefits arising from its provision. There are a number of approaches for doing so, but cost-utility analysis is favored by Ireland’s National Centre for Pharmacoeconomics [1], UK’s National Institute for Health and Care Excellence [2], and elsewhere. Utility can be defined in a number of ways, but it is commonly measured for these purposes in incremental “quality-adjusted life-years,” a combination of cumulative improvements in length of life and health-related quality of life likely to be achieved by the population if the service is provided.

One approach for measuring health-related quality of life is the three-level EuroQol five-dimensional questionnaire, as developed by EuroQol [3]. The technique measures five dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) on three levels (no problems, moderate problems, and severe problems), thus representing 243 (3^5) potential health states (HSs); unconsciousness and death are also included. By convention, full health is represented by a utility of 1 and dead is represented by a utility of 0. Some HSs may be considered worse than dead (WTD), and are given a negative utility.

This article explores how the analytic hierarchy process (AHP), as described by Saaty [4], might offer an appropriate framework to allow participants to assign utilities to HSs using ordinal valuation methods. AHP has been used in various health care settings [5]. The approach is tested using the data generated by the Measurement and Valuation of Health (MVH) study carried out in the United Kingdom in 1993 [6].

Conflict of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Brian P. Reddy, St. James Hospital, National Centre for Pharmacoeconomics, James Street, Dublin 8, Ireland.
E-mail: reddybr@tcd.ie.

1098-3015$36.00 – see front matter Copyright © 2015, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).
Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2015.05.010

The MVH study surveyed about 3000 members of the general public randomly selected as participants [6]. Forty-five HSs (including full health, unconsciousness, and death) were investigated, with 13 of these scored by each participant, ranging from "very mild" to "severe problems." HSs under consideration were first placed in order from best (1st) to worst (13th) by participants (referred to henceforth as *health state ranking* [HSR]), followed by cardinal ratings using the visual analogue scale (VAS) and time trade-off (TTO) approaches. Participants carried out VAS scoring directly after the HSR, and the approach required participants to give each HS a score from 0 (worst imaginable HS) to 100 (best imaginable HS).

The TTO approach asked patients to choose between living for 10 years in a given HS, or fewer years in full health. Utilities could be derived from their choices, and subsequently published on the basis of results of this method. For states considered WTD, various approaches have been used, but the MVH survey asked participants how many years of subsequent good health would be necessary to balance time in a given HS.

The TTO approach used in the MVH study allowed participants to value WTD HSs with utilities theoretically as low as –39 (where 3 months of poor health would be balanced only by 9 years and 9 months of subsequent full health). The authors of the original study considered that such results were unrealistic and artifactual. Utilities assigned for HSs were therefore bound between +1 and –1 before the mean utility for each mean HS was calculated. This method therefore requires some information (<–1) to be discarded, relies on the fallible and potentially arbitrary judgment of researchers, and inevitably removes potentially useful preference information. This approach has been acknowledged as imperfect, and other techniques have been tested [7] though none has been universally accepted as having solved the underlying issue. The AHP approach outlined in this article does not discard the information in this way and instead ultimately converts all information into scores, which might be considered analogous to utilities. We propose a simple method to derive utilities from these.

This article offers a preliminary analysis testing possible new approaches to examine population's HS preferences. The AHP approach outlined is derived explicitly from decision theory. It structured the process of transforming ordinal preference information into potentially meaningful utilities. The potential to derive utilities from ordinal data is what distinguishes this approach from similar approaches previously carried out investigating pairwise comparisons of the data, including an approach outlined in the original MVH report [8].

Theoretically, such an approach could allow participants to simply rank HSs in terms of preference and, given sufficient numbers of participants, conclusions could be drawn regarding the relative performance of each in terms of utilities. Estimates of the utility of HSs not ordered directly might be possible using the standard econometric approaches.

There are other relevant reasons to investigate the AHP approach. The original MVH study applied Thurstone's [9] Law of Comparative Judgment to investigate the strength of pairwise comparisons, which bears a clear resemblance to AHP [10]. AHP is one of a variety of multiple criteria decision analysis techniques available designed to help structure complex problems. It also seems timely, in a climate in which such techniques are increasingly being advocated for use in national health services [11], to investigate whether AHP might be useful in drawing meaningful conclusions from the data in determining HS preferences.

The aim of this study was therefore to investigate whether an approach based on AHP can be used to calculate meaningful utilities on the basis of analysis of pairwise comparisons of the ordinal preference data in a national EuroQol five-dimensional questionnaire survey.

Methods

AHP allows decision makers to build up a numerical score, based on their preferences derived from pairwise comparisons. When comparing how well each alternative under consideration has performed, these are reflected in an ordinal scale, derived by Saaty [4], which reflects the magnitude of how well each has done in qualitative, easy-to-understand statements, which are converted to a 1 to 9 numerical scale, as presented in Table 1. Over time, all such pairwise comparisons can be carried out and subsequently analyzed, and scores for each derived.

In this study, a similar protocol that compared the preferences of the general public, two HSs at a time, was followed. Doing this for every possible pair of HSs allowed scores for each to be generated. Utilities could subsequently be derived.

This scale has been used elsewhere as part of AHP analyses to allow qualitative descriptions of criteria derived from Delphi-style processes to be translated into a numerical scale, and subsequently analyzed [4,12,13]. In such cases, the importance of criteria and the subsequent performance of alternative courses of action on these criteria can be assessed. These can ultimately be combined to give each alternative course of action a unique overall score. AHP is therefore normally considered a multistage process, but for this study, only one such stage is required, comparing the proportion of participants who preferred each HS. The AHP approach is used as a framework by which to combine the matrix of ordinal relationships into normalized scores, indicating the relative performance of each HS.

In the MVH study, each participant ranked 13 semi-randomly selected HSs using the three previously described approaches. These states were considered simultaneously for HSR, and one at a time for both VAS and TTO. These can be considered as pairwise comparisons by examining how often one HS is preferred over another.

For any two states, the number of times they were compared by the same participant was measured, and subsequently on what proportion of occasions each HS was "preferred" ordinally. For HSR, this meant whichever HS was ranked higher by the

Table 1 – Scale derived by Saaty [4], used to convert pairwise qualitative relationships into a cardinal scale.

Intensity of weight	Definition
1	Equal importance
3	Weak moderate importance of one over another
5	Essential or strong importance
7	Very strong or demonstrated importance
9	Absolute importance
2, 4, 6, 8	Intermediate values between two adjacent scales
Reciprocals of above nonzero number	If activity *i* has one of the above nonzero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*

Note. The analytic hierarchy process uses a Likert-type scale following the template explained above. A number between 1 and 9 represents the more important/preferred of the pair, and the less preferred is given the reciprocal of this number. Comparisons of the public's preferences between health states were converted into this scale to generate scores for each health state, and subsequently to calculate utilities.

Table 2 – Four scales, and sample results, used to translate pairwise proportions of participant preference for each health state into Saaty's scale, described in Table 1.

	Formula used for $x \geq 0.5$	0	0.25	0.5	0.75	1
Linear approach	$1 + 16 \times (x - 0.5)$	1/9	1/5	1	5	9
Power	$9\text{^}(2 \times (x - 0.5))$	1/9	1/3	1	3	9
Beta distribution	$1 + 8 \times \beta \times (x - 0.5\|1,9)$	1/8.98	1/8.40	1	8.4	8.98
Logit	$\min(1 + \log[x/(1 - x)],9)$	1/9	1/1.48	1	1.48	9
Fitted beta	$1 + 8 \times \beta\ (x - 0.5\|8.0418,0.0001)$	1/1.0000000701	1/1.0000000002	1	1.0000000002	1.0000000701

Note. Algorithms tested to convert the proportion of participants who preferred each health state in a pairwise comparison into Saaty's scale. If an $\times$ is preferred by 75% of the participants (and the other therefore preferred by 25%), then using the linear algorithm above its associated value on Saaty's scale is 5 (and the other health state given the reciprocal, 1/5). These are sample values, and each algorithm is continuous, which are shown in Figure 1.

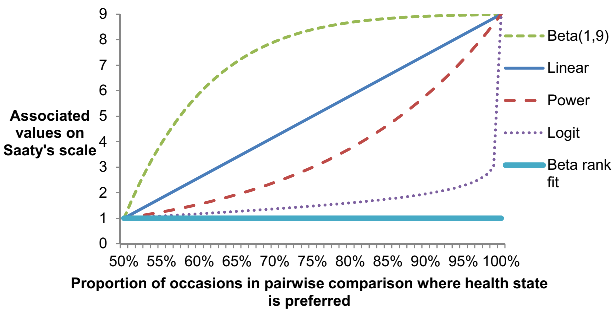

Fig. 1 – Continuous AHP algorithmic distributions used for health states preferred more than 50% of the time in head-to-head comparisons. States preferred less than 50% of the time are given the reciprocal of this figure. AHP, analytic hierarchy process. (Color version of figure is available online.)

P	A	B	C
A	NA	56.25	93.75
B	43.75	NA	62.5
C	6.25	37.5	NA

Q	A	B	C
	1.00	2	8
	0.5	1.00	3
	0.125	0.333	1.00

R	A	B	C
	0.615	0.600	0.667
	0.307	0.300	0.250
	0.077	0.100	0.083

S	Score
A	0.627
B	0.286
C	0.087

Fig. 2 – Sample figures for 3 × 3 matrix—P: Head-to-head comparisons of preference (e.g., 56.25% of the participants preferred A to B, and 43.75% preferred B to A); Q: Converted into the AHP scale, using linear algorithm in Table 2 (e.g., 1 + 16 × [0.5625 – 0.5] = 2, and hence for the reciprocal, ½ = 0.5); R: Normalized matrix using this scale (e.g., 1/[1 + 0.5 + 0.125] = 0.615); S: Score derived for each health state (e.g., [0.615 + 0.6 + 0.667]/3 = 0.627). AHP, analytic hierarchy process.

respondent. For VAS, whichever HS was given a higher score was preferred, and for TTO, this was for the HS given higher utility. Each was carried out for all 45 HSs. For each of these approaches, it can be represented by a 45 $\times$ 45 matrix, thus representing all such relationships. A score for each HS can be derived on the basis of these relationships, independently for each of HSR, VAS, and TTO.

However, there is no standard approach by which to transform the given proportions preferred in pairwise comparisons into Saaty's scale. Therefore, a number of alternative algorithms for doing so were tested, sample values of which are presented in Table 2. In all cases, it was assumed that if two HSs were preferred to each other on an equal number of occasions (i.e., 50% of the time), this would translate to "1" on this scale for both, indicating equal performance. Where possible, the whole scale was used, stretching from 1/9 to 9. Continuous distributions of each algorithm (for the preferred HS) are shown in Figure 1. For HSs preferred less than 50% of the time, the reciprocal of the corresponding figure above 50% is used (hence, for the linear approach, 75% of the people preferring an HS corresponds to 5 on Saaty's scale [$1 + 16 \times (0.75 - 0.5)$], and therefore 25% corresponds to 1/5).

For each of HSR, VAS, and TTO, a 45 $\times$ 45 matrix could be created using each of the distributions in Table 2, in turn, showing the pairwise relationships between HSs. Scores can be calculated from this by deriving the matrix's "maximal eigenvector," explained in Figure 2. The highest score represents the most preferred state, and HSs can subsequently be ranked. Theoretically, these scores could be transformed into utilities for each HS by fixing the values for full health (11111) at 1 and dead at 0 and calculating other values in proportion to these using a fixed scale.

A simpler example of 3 HSs (A, B, and C) is described in Figure 2. Matrix *P* shows the proportion of participants preferring each HS for each pairwise comparison (and hence entries diagonally opposite from each other sum to 100%). This is converted using the linear algorithm described above, and shown in *Q*. The subsequent diagonal in *Q* is composed of ones because each health state is assumed to be of equal performance to itself. Matrix *R* represents the normalized score of each HS within each column. *S* gives the score for each HS, equal to the mean of each row of *R*. Utilities can thereafter be calculated. If in this case HS A represented full health and B dead, C would represent an HS of (0.286 – 0.087)/(0.627 – 0.087), equal to 0.368.

An important stage following from this is to ensure that the completed matrix is then checked for consistency to ensure that meaningful conclusions can be derived, which was carried out for each of HSR, VAS, and TTO, using the conventional approach [4].

Results

This previously described process was carried out separately for HSR, VAS, and TTO approaches, and could feasibly be used for estimating each HS's overall rank and subsequently for estimating its utility. All results were similar, so the results explained here will focus on those of the HSR exercise. This was the only approach to elicit preferences from participants using an ordinal approach, making it the simplest of the three approaches. Because it requires less analytical thought, it may be more prone to inconsistent responses and hence lead to the most difficulty in modeling. The approach also perhaps holds the greatest promise for the AHP technique; if it can be shown to be effective at predicting TTO utilities in these circumstances, it may lead to a number of opportunities.

Figure 3 shows clear correlation in ranking derived for each of the 45 HSs from the AHP approaches (using each of the algorithms) and the rankings found in the MVH article. One obvious

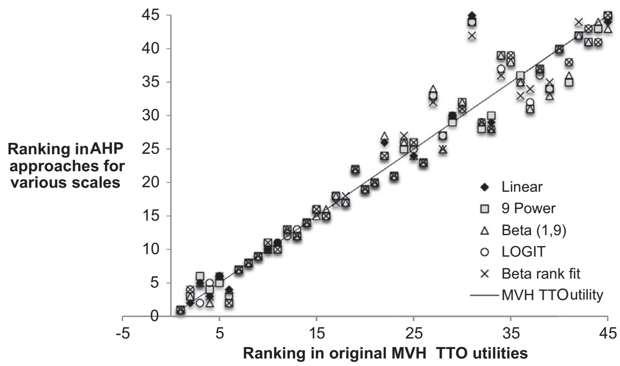

Fig. 3 – Comparisons showing similarities of best to worst health state rankings for MVH TTO-derived utilities and various AHP-derived scores, using data generated from participants using HSR. AHP, analytic hierarchy process; HSR, health state ranking; MVH, Measurement and Valuation of Health; TTO, time trade-off.

anomaly is the ranking of death—it is the lowest (or among the lowest) ranked of the AHP approaches, but was ranked about 30/45 for MVH TTO-derived utilities. This reflects differences between the results of TTO and HSR (and VAS) approaches. Subsequently, about one-third of the MVH TTO health states are considered WTD, whereas no states were considered WTD for either VAS or HS ranking. This phenomenon has been previously described in the original study [8], and could be seen in the data set based on the HS means and medians before carrying out the analysis. Ranking between the algorithms varies slightly but showed very similar results overall.

Although more parsimonious techniques such as average ranking of each state using the HSR approach would provide us with similar results more easily, the AHP approach has the advantage of subsequently generating scores that can be used to derive utilities. The HSR approach describes directly measured ordinal relationships, and when compared against the scores derived from the MVH TTO study (the black diagonal line in Figure 4), most AHP-derived utilities display a characteristic arching shape. This reflects the bunched nature of the results, with states considered WTD in the MVH TTO utilities rated better than expected (indeed few were considered WTD using this approach) and most other HSs rated lower. This occurred similarly for VAS and TTO approaches. For the *fitted beta* algorithm, these parameters were optimized to minimize the sum of squared errors between the resulting distribution of TTO-derived utilities and the MVH utilities. This fitted the distribution more closely, implying that it may be possible to predict results of a more thorough TTO study using a ranking approach alone.

Although derived from the same data, the AHP algorithms used had varying degrees of consistency. The beta rank fit [CR $<<$ 0.01], and Logit [0.027] were well below the 10% threshold, whereas power [0.19], linear [0.31], and beta (1,9) [0.55] were found to be inconsistent.

Discussion

This article intended to investigate whether AHP might be used to transform ordinal preferences into meaningful utilities. It appears that it may be possible to do so. Although no “single correct” answer can be derived from aggregated preference rankings [14], suitable functions can be found that appear to suitably mimic the TTO findings.

The approach outlined is similar to that of a discrete choice experiment, which is also based on pairwise comparisons and from which we can draw links to Thurstone’s [9] Law of Comparative Judgment. The key difference from the participants’ perspective is that for HSR they are asked to rank multiple (in this case 13) HSs simultaneously, which are subsequently analyzed, rather than building up a picture on the basis of comparisons of alternatives with varying performances on multiple criteria.

From a research point of view, this article highlights the broad family of utilities that may be derived from the same set of data. Given the differences found between TTO and VAS—and elsewhere with standard gamble [15]—it is not necessarily clear which approach is the “gold standard.” In its favor, AHP appears to offer a simple approach by which to estimate the results of more complex methods. Such an approach could feasibly address issues related to investigating population preferences in resource-limited settings [6], such as for subpopulations within

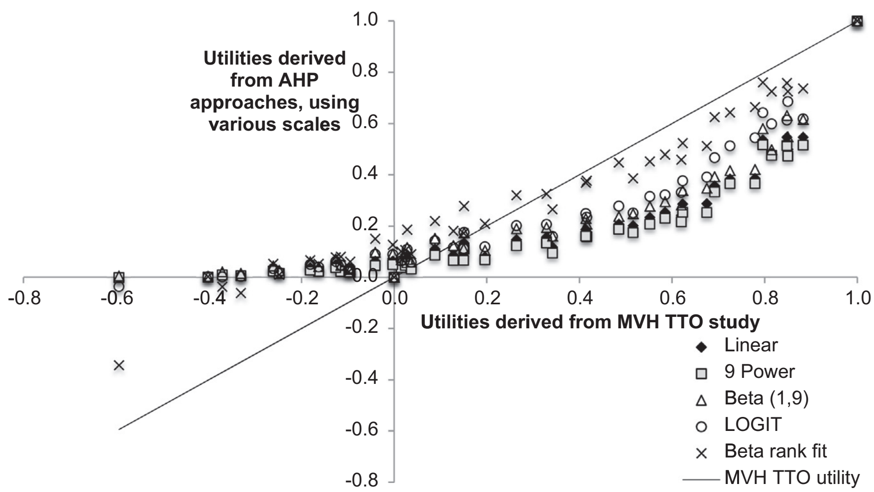

Fig. 4 – Derived utilities for each of the algorithms used, in comparison to the utilities derived using TTO in the MVH study. This was carried out using the HSR methodology. Note the characteristic curve, which was also present when the AHP technique was applied to VAS and TTO data. Beta rank fit best matches the utilities originally described in the MVH study, and was also found to be the most consistent technique. AHP, analytic hierarchy process; HSR, health state ranking; MVH, Measurement and Valuation of Health; TTO, time trade-off; VAS, visual analogue scale.

the United Kingdom, or other societies abroad. Because it is less cognitively demanding, it could potentially be carried out more quickly and cheaply than is possible at present, such as as a part of broader health surveys.

Results derived from the approach appear to be less extreme in WTD HSs than with the original TTO-derived utilities. This too would be a useful finding if corroborated elsewhere. They also have the advantage of having included all relevant preference data, rather than using arbitrary cutoffs. However, the final results are nonetheless derived by optimizing against the results of the TTO approach, and therefore these cutoffs in some sense remain indirectly. It would be interesting to further investigate how this issue can be mitigated in future.

The fact that findings derived from the same data may not vary only by accuracy but also by consistency was initially surprising. Whether the 10% threshold for consistency is appropriate for population-level surveys (as opposed to more manageable, boardroom-sized groups seen in the previous Delphi-style examples) is perhaps debatable. It is possible, for example, that two or more subgroups in such a larger population may express preferences that, when aggregated, lead to inconsistent or seemingly irrational preferences. However, given the availability of algorithms found to produce consistent rankings, it seems prudent to use these and reject others. This issue does highlight that under certain circumstances in future, it is conceivable that decision makers may have to trade off the accuracy of the model against consistency. How to negotiate such a situation is beyond the scope of this article.

The initial assumption that the best fit would most likely use the full range of Saaty's scale, from 1/9 to 9, was also wrong. The best fit counterintuitively barely moved at all around 1. Further testing showed that if we were to use an algorithm that assumed all pairwise comparisons were equal, deriving utilities was infeasible. Nonetheless, the narrow range found in the optimized fit was surprising.

There are a number of limitations to the findings, most obviously over whether the approach is subject to overfitting, and whether the parameters derived would be meaningful for another participant population such as the general public of another society, or subpopulations within the United Kingdom. This will require further research in future if the approach is to be used in practice. The article has not investigated the minimum number of participants required to produce reliable utilities. In practice, simpler algorithmic approaches may also be possible and preferable. The approach used to convert AHP scores into utilities may also have other transformations possible, further complicating the optimization stage.

Inconsistent responses, where participants' ranking of HSs changed depending on the approach used, were not removed from the data set and all records were therefore included in the analyses. Various approaches have been proposed for dealing with such inconsistences in the past, but given the nature of ordinal population-level pairwise relationships, it was felt that decisions about whether to exclude outliers would have limited impact anyway. Seeing that calculations for HSR, VAS, and TTO were carried out independently, this will not lead to within-approach inconsistencies, but may allow a small number of implausible preferences to be included.

Standard gamble and discrete choice experiment were not used in the initial survey, so these could not be tested using an AHP approach or their performance compared against those of other techniques. It would be interesting to investigate these further in future.

Further research is required if this approach were to be used in practice. It is as yet unclear whether the parameters used in this case would lead to useful predictions for another population. If not, it is unclear how we might decide which parameters could be used when there is no "correct" TTO-derived utilities available against which to optimize the approach. It may prove that some parameters are universal, or universal enough to be practicable, but there is not yet any basis to conclude this.

This article described an approach that may be used to convert ordinal preference data from sufficient numbers of participants into HS utilities. It shows that such an approach is possible though it cannot provide definitive answers to all relevant questions that arise from this. In time, it is hoped that further research, testing the approach in other populations, will shed some light on some of these issues [16].

Source of financial support: This work was funded primarily by the Health Research Board through Health Research Award HRA-2014-PHR-516. Further funding for Prof. Walsh was provided through grant number HRB-RL-2013-4.

REFERENCES

[1] Barry M, Usher C, Tilson L. Public drug expenditure in the Republic of Ireland. Expert Rev Pharmacoecon Outcomes Res 2010;10:239–45.
[2] Kelly M, Morgan A, Ellis S, et al. Evidence based public health: a review of the experience of the National Institute of Health and Clinical Excellence (NICE) of developing public health guidance in England. Soc Sci Med 2010;71:1056–62.
[3] Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53–72.
[4] Saaty TL. How to make a decision: the analytic hierarchy process. Eur J Oper Res 1990;48:9–26.
[5] Liberatore MJ, Nydick RL. The analytic hierarchy process in medical and health care decision making: a literature review. Eur J Oper Res 2007;189:194–207.
[6] Williams A. The Measurement and Valuation of Health: A Chronicle. Discussion paper 136, center for Health Economics, York, 1995.
[7] Devlin N, Buckingham K, Shah K, et al. A comparison of alternative variants of the lead and lag time TTO. Health Econ 2013;22:517–32.
[8] Dolan P, Guldex Cl, Kind P, Williams A. The Measurement and Valuation of Health: First Report on the Main Survey. York: Centre for Health Economics, University of York, 1994.
[9] Thurstone LL. A law of comparative judgment. Psychol Rev 1927;34:273.
[10] MacKay DB, Bowen WM, Zinnes JL. A Thurstonian view of the analytic hierarchy process. Eur J Oper Res 1996;89:427–44.
[11] Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care: a bibliometric analysis. Oper Res Health Care 2013;2:20–4.
[12] Dolan JG. Involving patients in decisions regarding preventive health interventions using the analytic hierarchy process. Health Expect 2000;3:37–45.
[13] Reddy BP, Kelly MP, Thokala P, et al. Prioritising public health guidance topics in the National Institute for Health and Care Excellence using the Analytic Hierarchy Process. Public Health 2014;128:896–903.
[14] Arrow KJ. A difficulty in the concept of social welfare. J Polit Econ 1950;58:328–46.
[15] Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Econ 2002;11:447–56.
[16] Donegan HA, Dodd FJ. A note on Saaty's random indexes. Math Comput Modell 1991;15:135–7.